

Diversity scaling of human vaginal microbial communities

DEAR EDITOR,

The composition and diversity of the human vaginal microbial community have been investigated intensively due to the diversity-stability relationship (DSR)-based hypothesis for bacterial vaginosis (BV) etiology, which was first proposed in the 1990s and has received renewed interest in recent years. Nevertheless, diversity changes (scaling) across individuals in a cohort or population have not yet been addressed, which is significant both theoretically and practically. Theoretically, biodiversity scaling is the core of biogeography, and practically, inter-subject heterogeneity is critical for understanding the etiology and epidemiology of human microbiome-associated diseases such as BV. Here we applied the diversity-area relationship (DAR), a recent extension to the classic species-area relationship (SAR), to study diversity scaling of the vaginal microbiome by reanalyzing reported data collected from 1107 postpartum women. The model used here characterized the power-law (or its extension) relationships between accrued diversity and areas (numbers of individuals), upon which four biogeographic profiles were thus defined. Specifically, we established the DAR profile (relationship between diversity scaling parameter and so-called diversity order (q)), similarly pair-wise diversity overlap (PDO) profile, maximal accrual diversity (MAD) profile, and ratio of individual-level to population-level diversity (RIP) profile. These four profiles offer valuable tools to assess and predict diversity scaling (changes) in the human vaginal microbiome across individuals, as well as to understand the dynamics of vaginal microbiomes in healthy women.

The human vaginal microbiome is a complex ecosystem that plays critical roles in maintaining host health. As the first defense of the reproductive tract, the vaginal microbiome is critical for the prevention of opportunistic pathogen colonization and viral infection. For example, endogenous,

healthy vaginal microbiota can help protect against HIV infection by activating local and systemic inflammation; however, microbiota associated with BV can also increase susceptibility to HIV infection (Buvé et al., 2014; Petrova et al., 2013). For pregnant women, the vaginal microbiota is not only associated with maternal health but also that of neonates, with the composition of the newly colonized microbiome playing a key role in newborn immunity and metabolic development (Cox et al., 2014; Dominguez-Bello et al., 2010; Olszak et al., 2012; Rutayisire et al., 2016). Furthermore, babies delivered by cesarean section can have a higher risk of metabolic and immune diseases than those delivered vaginally (Dominguez-Bello et al., 2010; Sevelsted et al., 2015; Younes et al., 2018), although Chu et al. (2017) noted that delivery mode does not influence microbiome composition in newborns. Moreover, in pregnancy, vaginal dysbiosis is hypothesized to be a contributor to spontaneous preterm birth (Freitas et al., 2018; Romero et al., 2014a; Stout et al., 2017) and miscarriage (Ralph et al., 1999).

In many healthy women, the vaginal microbiota is dominated by *Lactobacillus* spp. (Macklaim et al., 2013; Ravel et al., 2011). Several studies (Brotman et al., 2014; Gajer et al., 2012; Ma & Li, 2017; Ravel et al., 2011) have confirmed the five major community state types of the vaginal microbiome in adult women, as first identified by Ravel et al. (2011). Four types are dominated by *Lactobacillus* spp., including *L. iners*, *L. crispatus*, *L. gasseri*, and *L. jensenii*. However, 20%–30% of asymptomatic, otherwise healthy individuals lack lactic acid bacteria in their vaginal microbiome, which instead consists of obligate anaerobic bacteria (Ravel et al., 2011, 2013). In addition, the frequency of microbiome type varies in different ethnic groups, with those microbiome not dominated by *Lactobacillus* spp. more commonly found in healthy Hispanic and black women than in Asian or white women (Ma et al., 2012; Ravel et al., 2011). Furthermore, the composition of the vaginal microbiome is dynamic during life and associated with menopause stage (Muhleisen & Herbst-Karlovitz, 2016). Recent research demonstrated the vaginal microbiome of perimenarcheal adolescents to be dominated by *Lactobacillus* spp., including

Open Access

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2019 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 10 July 2019; Accepted: 09 September 2019; Online: 24 September 2019
DOI: 10.24272/j.issn.2095-8137.2019.068

L. crispatus, *L. iners*, *L. gasseri*, and *L. jensenii*, similar to that found in reproductive-age women (Hickey et al., 2015). In premenopausal women, the vaginal microbiota is still dominated by *L. crispatus* and *L. iners*, but *Lactobacillus* spp. are often replaced by *Streptococcus* and *Prevotella* in the perimenopausal and postmenopausal stages (Brotman et al., 2014). Shifts in vaginal microbiome have also been observed during and after pregnancy. For example, diversity and richness of the vaginal microbiome is lower in pregnant women than in non-pregnant women (Freitas et al., 2017). Furthermore, Romero et al. (2014b) showed that the vaginal microbiome of pregnant women contains a higher abundance of *L. vaginalis*, *L. crispatus*, *L. gasseri*, and *L. jensenii*, and a lower probability of switching to a *Lactobacillus*-deficient community. In addition, radical changes in *Lactobacillus*-poor vaginal communities have been found at delivery, which can persist for up to a year (DiGuilio et al., 2015).

Despite extensive studies on the human vaginal microbiome, what constitutes normal or healthy vaginal microbiota remains unresolved. For example, Doyle et al. (2018) sampled and sequenced the vaginal microbiome of 1 107 rural Malawi women after pregnancy, and found that 75.7% (752/994) of the population were dominated by *Gardnerella vaginalis* rather than by *Lactobacillus* spp., and although *L. iners* increased with time after delivery, *G. vaginalis* still dominated for an extended period. In Doyle's study, both the pregnancy delivery mode and ethnicity also appeared to influence the composition of vaginal microbiome, though all hosts were healthy.

Previous research has revealed that the biodiversity of vaginal microbial communities varies with health status and lifestyle of the host. Nevertheless, existing studies have not addressed diversity scaling (changes) across individuals in a cohort or population. Theoretically, microbiome diversity distribution across individual subjects (i. e., space) is traditionally a focus of microbial biogeography. Practically speaking, understanding the biogeography of the human microbiome can reveal critical information on its characteristics in a cohort setting, which can, in turn, significantly influence studies on the etiology and epidemiology of human microbiome-associated diseases such as inflammatory bowel disease, obesity, and BV. To effectively assess the spatial scaling of human vaginal microbial diversity, we applied the DAR model, which is a recent extension of the classic SAR in biogeography and conservation biology (Bell et al., 2005; Horner-Devin et al., 2004; MacArthur & Wilson, 1967; Noguez et al., 2005; Peay et al., 2007; Triantis et al., 2012; Várbiró et al., 2017; Whittaker & Triantis, 2012). SAR is one of the oldest described ecological laws or patterns, whereby species richness increases with increasing sampling area, and can be traced back to the 19th century (Watson, 1835). It is still considered one of the most important principles in conservation biology and biogeography. The extensions from SAR to DAR introduced a several important advances including: (1) Expanding species richness (number of species) to general diversity measures in Hill numbers (Chao et al., 2012,

2014; Hill, 1973), thus making it possible to not only assess the scaling of species richness (numbers), but also scaling of general diversity (e. g., change in community evenness or dominance). Therefore, the classic SAR is a special case of the more general DAR; (2) The DAR, PDO, MAD, and local regional/global diversity (LRD/LGD) profiles are effective tools for the biogeographic mapping of biodiversity over space (Ma, 2018a, 2018c, 2019).

In this study, we applied DAR modeling and associated biogeographic profiles to investigate the spatial diversity scaling of postpartum vaginal microbial communities across individuals by reanalyzing the large vaginal microbiome dataset originally reported by Doyle et al. (2018). The spatial diversity scaling of the vaginal microbiome revealed heterogeneity among individuals, which could provide an ecological basis for personalized and precise diagnosis and treatment of microbiome-associated diseases, including BV. The biogeographic profiles of the vaginal microbiome also provide tools for explaining the DSR hypothesis for BV etiology from multiple dimensions (Ma & Ellison, 2018, 2019).

The vaginal microbial dataset (Doyle et al., 2018) reanalyzed in this study consisted of 1 158 vaginal microbiome samples collected from 1 107 rural Malawi women post-delivery. Most samples were collected within the first 20 d of delivery, though some were sampled 5–583 d post-delivery. The V5–V7 hypervariable regions of the 16S rRNA genes were amplified and sequenced under the MiSeq Illumina platform. After quality control, the sequences were clustered into 14 354 operational taxonomic units (OTUs) using QIIME 2.8.6. Samples with less than 2 000 reads were removed, as were OTUs with less than 1 000 reads. After prescreening, 1 076 samples and 466 OTUs remained for DAR analysis. In DAR analysis, the number of each OTU read is equivalent to the population abundance of a species in macro-ecology, or OTU abundance in diversity analysis. More detailed information on the dataset can be found in Doyle et al. (2018).

The Hill numbers (Hill, 1973) were reintroduced to ecology by Jost (2007) and Chao et al. (2012, 2014), and possess certain critical advantages over traditional diversity indexes. The Hill numbers for measuring alpha diversity are as follows:

$$^qD = \left(\sum_{i=1}^S p_i^q \right)^{1/(1-q)} \quad (1)$$

When $q=1$, the Hill number is undefined, but its limit exists in the following form:

$$^1D = \lim_{q \rightarrow 1} ^qD = \exp \left(- \sum_{i=1}^S p_i \log(p_i) \right) \quad (2)$$

where, D is the diversity in Hill numbers, q ($=0, 1, 2, \dots$) is the order number of diversity, S is the number of species (or OTUs), and p_i is the relative abundance of species i . The diversity order (q) sets the sensitivity of the Hill numbers to the relative frequencies of species abundances. When $q=0$, 0D is equal to the number of species or species richness (S). When $q=1$, 1D is the number of typical or common species in the community and is equal to the exponential of Shannon entropy. When $q=2$, 2D is more sensitive to species with high abundance, and is equal to the inverse of the Simpson index.

Generally, qD is the diversity of a community with $x={}^qD$ equally abundant species.

Beta-diversity can be defined with the multiplicative partitioning of Hill numbers (Chao et al., 2012, 2014; Ellison, 2010; Gotelli & Chao, 2013; Jost, 2007), as follows:

$${}^qD_{\beta} = {}^qD_{\gamma} / {}^qD_{\alpha} \quad (3)$$

where, ${}^qD_{\alpha}$ and ${}^qD_{\gamma}$ are the alpha and gamma diversities in terms of Hill numbers, respectively. As ${}^qD_{\gamma}$ is equivalent to the alpha diversity of the meta-community, it has the same definition as alpha-diversity (Eqn. (1)). Chao et al. (2012, 2014) defined a series of Hill numbers corresponding to different diversity orders (q) as the diversity profile. In this study, the diversity or Hill numbers were computed until the third order, $q=3$.

According to Ma (2018a), we used the power law (PL) DAR model and power-law with exponential cutoff (PLEC) model as the DAR models for the human vaginal microbiome. The PL model is:

$${}^qD = cA^z \quad (4)$$

where, qD is diversity measured in Hill numbers of the q -th order, A is the area (number of individuals), and c & z are the PL parameters.

The PLEC model is:

$${}^qD = cA^z \exp(dA), \quad (5)$$

where, d is a third parameter that is usually less than zero in DAR modeling, and $\exp(dA)$ is then the exponential decay item that eventually overwhelms the power law behavior when A is sufficiently large.

To simplify parameter estimation, we transformed non-linear Equations (4) and (5) into log-linear regression equations:

$$\ln(D) = \ln(c) + z\ln(A) \quad (6)$$

$$\ln(D) = \ln(c) + z\ln(A) + dA \quad (7)$$

In Eqn. (6), z is the slope of the log-linear transformed PL model, which is equivalent to its interpretation in the traditional SAR—ratio of diversity accrual rate to area increase rate. Parameter c of the PL model can be viewed as the number of species equivalent to diversity in the first unit of area to accrue. Thus, the accrual order of area unit may influence parameter c . To deal with this technical issue, the units (individuals/samples) to be accumulated were randomly permuted each time the DAR model was built. For each dataset, we repeatedly applied DAR modeling 100 times by randomly re-ordering all samples in the dataset. For the detailed computational procedure, please refer to Ma (2018a).

Similar to the diversity profile concept of Chao et al. (2012, 2014), which is a series of Hill numbers corresponding to different diversity orders (q), Ma (2018a) and Ma & Li (2018) proposed four DAR-based profiles, including the DAR, PDO, MAD, and LRD/LGD profiles. These four profiles can be quantitatively characterized by parameters from the PL/PLEC DAR models and can be used to sketch out biogeography maps of the human microbiome or other ecological communities.

The DAR profile was defined as a series of z -values (scaling parameter) of the PL-DAR model (Eqns. 4 & 6), i.e., a series of z -values corresponding to different diversity orders (q) or z - q trends.

The PDO profile was defined as:

$$g = 2^{-2^z} \quad (8)$$

where, z is the scaling parameter of the PL-DAR model, i.e., the PDO profile is a series of g (q) values corresponding to different diversity orders (q), computed with Eqn. (8).

The MAD profile was defined as a series of MAD or ${}^qD_{\max}$ values, corresponding to different diversity orders (q):

$${}^qD_{\max} = c \left(-\frac{z}{d} \right)^z \exp(-z) = cA_{\max}^z \exp(-z) \quad (9)$$

where, $A_{\max} = -z/d$ is the number of individuals (samples) needed to reach the MAD, and c and z are parameters of the PLEC-DAR model (Eqns. (5) & (7)).

The RIP profile was defined as a series of RIP values corresponding to different diversity orders (q), as specified by the following equation:

$${}^qRIP = {}^qC / {}^qD \quad (10)$$

where, c is a parameter of the PL-DAR model and D is the diversity in Hill numbers estimated with the PLEC-DAR model (Eqns. (5) & (7)). Based on the above RIP definition, a RIP profile can be defined for a population (cohort) of any size. In practice, using ${}^qD_{\max}$ for qD is more convenient, i.e.:

$${}^qRIP = {}^qC / {}^qD_{\max}. \quad (11)$$

The RIP parameter assesses the average level of an individual to represent a population (or cohort) from which the individual is a member. The RIP profile is also known as the LRD (local-to-regional diversity) or LGD (local-to-global diversity) profile in other ecological systems beyond the human microbiome (Ma & Li, 2018; Ma, 2019).

We built two DAR models for the vaginal microbiome, including the PL and PLEC models for alpha-diversity and beta-diversity scaling, respectively. The results are listed in Tables 1 and 2, including the diversity order (q) of Hill numbers, mean model parameters (z , $\ln(c)$, d , g , D_{\max}) and their standard errors, and measures (correlation coefficient R & P -value) for goodness-of-fitting. N represented the number of successful fittings out of 100 p re-samplings, as explained previously. Re-sampling was performed to deal with the possible influence of the order of diversity accrual (i.e., order in which the samples were accrued for building the DAR model) on model parameter c . Except for two cases of alpha-DAR modeling at diversity order $q=3$, the fittings to the DAR models were successful in all 100 re-samplings. Even in the two exceptions, the success rates were 97% and 99%, respectively. Therefore, the DAR models were considered suitable for vaginal microbiome assessment, as also evident by the R (linear correlation coefficient) and associated p values, which indicated the goodness-of-fit of the DAR models.

Based on Table 1, we found the following in regard to alpha-DAR scaling:

(1) As one of the most important parameters from the PL-DAR model, the scaling parameter (z) at different diversity orders (q) was $z(q) = (0.807(0), 0.171(1), 0.110(2), 0.095(3))$, where $z(q)$ represents the DAR profile according to previous definition. The DAR profile characterizes the diversity scaling across individuals (over space) comprehensively. Results also showed that the scaling level differed at different orders. For

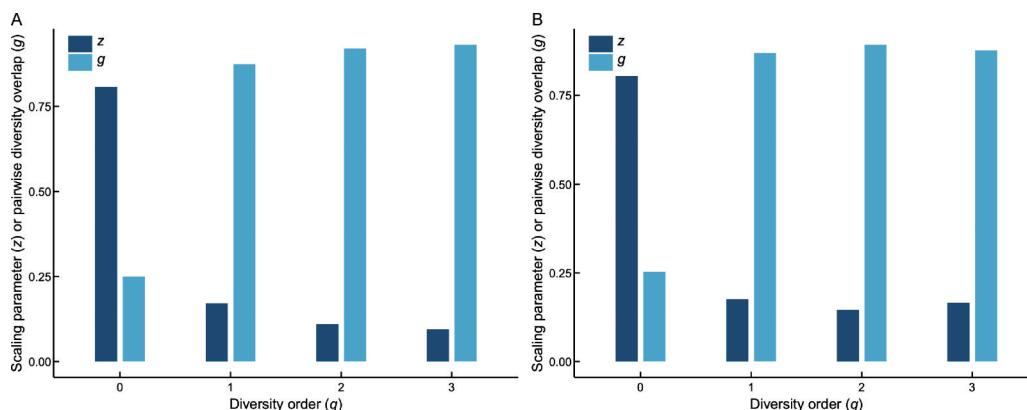
Table 1 Alpha-DAR models computed with 100 re-samplings for the vaginal microbiome

| Diversity order and statistics | Power law (PL) | | | | | | PL with exponential cutoff (PLEC) | | | | | | | |
|--------------------------------|----------------|--------|-------|---------|-------|-------|-----------------------------------|---------|-------|-------|---------|-----|------------------|------------------|
| | z | ln(c) | R | P-value | g | N | z | d | ln(c) | R | P-value | N | A _{max} | D _{max} |
| q=0 | Mean | 0.807 | 3.902 | 0.998 | 0.000 | 0.250 | 0.778 | 0.0001 | 4.019 | 0.999 | 0.000 | 100 | N/A | N/A |
| | SE | 0.031 | 0.200 | 0.001 | 0.000 | 0.038 | 0.062 | 0.0001 | 0.313 | 0.001 | 0.000 | | | |
| | Min | 0.738 | 3.474 | 0.995 | 0.000 | 0.160 | 0.642 | -0.0002 | 3.201 | 0.995 | 0.000 | | | |
| q=1 | Max | 0.880 | 4.356 | 1.000 | 0.000 | 0.332 | 0.948 | 0.0004 | 4.683 | 1.000 | 0.000 | 100 | 667 | 86.8 |
| | Mean | 0.171 | 3.345 | 0.860 | 0.000 | 0.874 | 0.291 | -0.0004 | 2.861 | 0.930 | 0.000 | | | |
| | SE | 0.039 | 0.257 | 0.073 | 0.000 | 0.031 | 0.074 | 0.0002 | 0.374 | 0.044 | 0.000 | | | |
| q=2 | Min | 0.085 | 2.669 | 0.582 | 0.000 | 0.789 | 0.083 | -0.0008 | 1.820 | 0.782 | 0.000 | 100 | 530 | 34.4 |
| | Max | 0.276 | 3.951 | 0.963 | 0.000 | 0.940 | 0.488 | 0.0001 | 3.831 | 0.990 | 0.000 | | | |
| | Mean | 0.110 | 2.808 | 0.681 | 0.000 | 0.920 | 0.229 | -0.0004 | 2.328 | 0.829 | 0.000 | | | |
| q=3 | SE | 0.047 | 0.312 | 0.173 | 0.000 | 0.035 | 0.080 | 0.0002 | 0.411 | 0.107 | 0.000 | 99 | 507 | 24.5 |
| | Min | 0.013 | 1.901 | 0.155 | 0.000 | 0.814 | 0.023 | -0.0008 | 1.162 | 0.443 | 0.000 | | | |
| | Max | 0.246 | 3.457 | 0.933 | 0.000 | 0.991 | 0.448 | 0.0002 | 3.225 | 0.974 | 0.000 | | | |
| | Mean | 0.095 | 2.565 | 0.589 | 0.013 | 0.931 | 0.209 | -0.0004 | 2.109 | 0.770 | 0.001 | 99 | 507 | 24.5 |
| | SE | 0.054 | 0.362 | 0.231 | 0.084 | 0.040 | 0.088 | 0.0002 | 0.447 | 0.157 | 0.014 | | | |
| | Min | -0.002 | 1.508 | 0.013 | 0.000 | 0.811 | -0.004 | -0.0009 | 0.890 | 0.061 | 0.000 | | | |
| | Max | 0.250 | 3.269 | 0.932 | 0.667 | 1.001 | 0.431 | 0.0003 | 3.112 | 0.971 | 0.138 | 99 | 507 | 24.5 |

z, *c* & *d*: The parameters of PL- and/or PLEC-DAR models, in which *z* of PL-DAR model is the scaling parameter, and *c* can be viewed as the number of species equivalent to diversity in the first unit of area to accrue. *g*: The pair-wise diversity overlap. *R*: The correlation coefficient to judge the goodness-of-fitting of PL or PLEC models. *P*-value: The parameter to judge the success or failure of the fitting of PL or PLEC models (*P*≤0.05 indicates successful model fitting). *N*: The number of successful fittings out of 100 re-samplings. *A_{max}*: The number of accrued individuals corresponding to the maximal accrual diversity. *D_{max}*: The maximal accrual diversity. *SE*: Standard error. *Min*: Minimum. *Max*: Maximum. *N/A*: Not available.

example, scaling at diversity order *q*=0, which is equivalent to the classic SAR law, was faster than that at *q*=1, 2, or 3, as

indicated by the monotonically decreasing *z*-value (see Figure 1A for alpha-DAR profile).

**Figure 1** DAR profile (*z*-*q*) and PDO profile (*g*-*q*) of the vaginal microbiome

A: Alpha-diversity scaling; B: Beta-diversity scaling.

(2) The PDO profile was $g(q)=(0.250(0), 0.874(1), 0.920(2), 0.931(3))$. The PDO profile, which characterizes the overlap or similarity between pair-wise individuals, showed the opposite trend as the DAR profile, i.e., a monotonically increasing trend (see Figure 1A for alpha-PDO profile).

(3) The MAD profile characterizes the theoretically maximal

accumulation of diversity across individuals. Here, regarding the MAD profile, the PLEC model failed to produce *D_{max}* at diversity order *q*=0 because *d*>0, for which a maximum does not exist. For the diversity orders *q*=1, 2, 3, the PLEC model for alpha-diversity successfully generated *D_{max}*, i.e., $D_{max}(q)=(86.8(1), 34.4(2), 24.5(3))$.

Table 2 Beta-DAR models computed with 100 re-samplings for the vaginal microbiome

| Diversity order and statistics | Power law (PL) | | | | | | PL with exponential cutoff (PLEC) | | | | | | | |
|--------------------------------|----------------|-------|--------|---------|-------|-------|-----------------------------------|----------|--------|-------|---------|-----|-----------|-----------|
| | z | ln(c) | R | P-value | g | N | z | d | ln(c) | R | P-value | N | A_{max} | D_{max} |
| $q=0$ | Mean | 0.805 | -0.070 | 1.000 | 0.000 | 0.253 | 0.770 | 0.00012 | 0.070 | 1.000 | 0.000 | 100 | N/A | N/A |
| | SE | 0.010 | 0.065 | 0.000 | 0.000 | 0.012 | 0.016 | 0.00004 | 0.078 | 0.000 | 0.000 | | | |
| | Min | 0.782 | -0.217 | 0.999 | 0.000 | 0.228 | 0.722 | -0.00002 | -0.131 | 1.000 | 0.000 | | | |
| | Max | 0.826 | 0.070 | 1.000 | 0.000 | 0.280 | 0.812 | 0.00022 | 0.313 | 1.000 | 0.000 | | | |
| $q=1$ | Mean | 0.176 | 1.591 | 0.884 | 0.000 | 0.870 | 0.298 | -0.00043 | 1.092 | 0.952 | 0.000 | 100 | 687 | 15.5 |
| | SE | 0.031 | 0.208 | 0.059 | 0.000 | 0.024 | 0.041 | 0.00011 | 0.219 | 0.034 | 0.000 | | | |
| | Min | 0.104 | 1.047 | 0.686 | 0.000 | 0.807 | 0.215 | -0.00076 | 0.449 | 0.829 | 0.000 | | | |
| | Max | 0.255 | 2.093 | 0.974 | 0.000 | 0.925 | 0.435 | -0.00022 | 1.570 | 0.994 | 0.000 | | | |
| $q=2$ | Mean | 0.146 | 1.890 | 0.734 | 0.000 | 0.893 | 0.291 | -0.00052 | 1.295 | 0.867 | 0.000 | 100 | 563 | 17.3 |
| | SE | 0.050 | 0.333 | 0.178 | 0.000 | 0.038 | 0.076 | 0.00024 | 0.368 | 0.111 | 0.000 | | | |
| | Min | 0.017 | 1.026 | 0.106 | 0.000 | 0.791 | 0.102 | -0.00124 | 0.227 | 0.335 | 0.000 | | | |
| | Max | 0.274 | 2.730 | 0.978 | 0.001 | 0.988 | 0.525 | -0.00002 | 2.356 | 0.987 | 0.000 | | | |
| $q=3$ | Mean | 0.166 | 1.963 | 0.711 | 0.000 | 0.877 | 0.330 | -0.00058 | 1.293 | 0.850 | 0.000 | 100 | 566 | 21.3 |
| | SE | 0.062 | 0.421 | 0.200 | 0.000 | 0.049 | 0.103 | 0.00035 | 0.468 | 0.129 | 0.000 | | | |
| | Min | 0.021 | 0.798 | 0.108 | 0.000 | 0.735 | 0.069 | -0.00149 | 0.115 | 0.203 | 0.000 | | | |
| | Max | 0.339 | 2.917 | 0.972 | 0.000 | 0.986 | 0.591 | 0.00016 | 2.663 | 0.980 | 0.000 | | | |

z , c & d : The parameters of PL- and/or PLEC-DAR models, in which z of PL-DAR model is the scaling parameter, and c can be viewed as the number of species equivalent to diversity in the first unit of area to accrue. g : The pair-wise diversity overlap. R : The correlation coefficient to judge the goodness-of-fitting of PL or PLEC models. P -value: The parameter to judge the success or failure of the fitting of PL or PLEC models ($P \leq 0.05$ indicates successful model fitting). N : The number of successful fittings out of 100 re-samplings. A_{max} : The number of accrued individuals corresponding to the maximal accrual diversity. D_{max} : The maximal accrual diversity. SE: Standard error. Min: Minimum. Max: Maximum. N/A: Not available.

From Table 2, we found the following in regard to beta-DAR scaling:

(1) As one of the most important parameters of the PL-DAR model, the beta-diversity scaling parameter (z) at different diversity orders (q) was $z(q) = (0.805(0), 0.176(1), 0.146(2), 0.166(3))$, where $z(q)$ represents the DAR profile according to previous definition and characterizes diversity scaling across individuals (over space) comprehensively (see Figure 1B for beta-DAR profile). Comparison between the beta-DAR and alpha-DAR profiles revealed an interesting phenomenon: i.e., the alpha-DAR profile monotonically decreased with q , whereas the beta-DAR profile was valley-shaped. This suggests that, at a lower diversity order (q), the alpha-DAR and beta-DAR scaling parameters (z) were rather close to each other, but the difference was enlarged at higher diversity orders (q).

(2) The beta-PDO profile was $g(q) = (0.253(0), 0.870(1), 0.893(2), 0.877(3))$ for beta-diversity scaling. Here, the PDO profile, which characterizes the overlap or similarity between pair-wise individuals, showed the opposite trend to the DAR profile, i.e., a bell-shaped trend (see Figure 1B for beta-PDO profile).

(3) The MAD profile characterizes the theoretical maximal accumulation of diversity across individuals. Here, regarding the beta-MAD profile, the PLEC model failed to produce D_{max} at diversity order $q=0$ because $d > 0$, for which a maximum

does not exist. For diversity orders $q=1, 2, 3$, the PLEC model for beta-diversity successfully generated D_{max} , i.e., $D_{max}(q) = (15.5(1), 17.3(2), 21.3(3))$.

Table 3 shows the RIP values for both alpha-DAR and beta-DAR of the vaginal microbiome. At diversity order $q=0$, the estimation of ${}^0D_{max}$ failed, and RIP for $q=0$ could not be estimated. For $q=1, 2, 3$, RIP was successfully estimated for alpha- and beta-diversity, respectively. Here, RIP characterized the relationship between individual- and population-level diversity. For example, at diversity order $q=1$, alpha-RIP=0.327 and beta-RIP=0.316, indicating that an average individual represented approximately 33% and 32% of population alpha- and beta-diversity, respectively.

In the current study, we investigated the diversity (including alpha- and beta-diversity) scaling of the human vaginal microbiome across individuals by re-analyzing a big dataset

Table 3 Ratio of individual-level to population-level diversity (RIP) of the vaginal microbiome

| Diversity order | RIP for alpha-DAR | RIP for beta-DAR |
|-----------------|-------------------|------------------|
| $q=0$ | N/A | N/A |
| $q=1$ | 0.327 | 0.316 |
| $q=2$ | 0.482 | 0.383 |
| $q=3$ | 0.530 | 0.335 |

N/A: Not available.

originally published by Doyle et al. (2018). Compared with the microbial SAR range reported in existing literature for other microbes, such as Green & Bohannan's (2006) range between 0.019–0.470, the scaling parameter (z) estimated in our study, i.e., α -z=0.807, β -z=0.805, appears to be out of the known range, at nearly twice that reported for SAR values for other microbes. Three possibilities exist for the significant difference: (1) The use of revolutionary metagenomic sequencing technology, which allows for detection of more microbial species and consequently large scaling parameter; (2) The human vaginal microbiome has higher heterogeneity across individuals, which could be validated by future biomedical studies; and (3) The postpartum nature of the vaginal microbiome samples analyzed in this study. We could not exclude these possibilities at present due to insufficient available data for comparative research. Indeed, previous studies have classified human vaginal microbiomes into five main community-state types (CSTs), in which CST I, II, III, and V are dominated by *Lactobacillus* spp., and CST IV is composed of facultative or strictly anaerobic bacteria (Gajer et al., 2012; Ravel et al., 2011), many of which are BV-related. The vaginal microbial communities of postpartum women in rural Malawi studied by Doyle et al. (2018) and reanalyzed here were mostly *Lactobacillus*-deficient microbiomes, which could be grouped as CST IV, although all these women were healthy. Therefore, the classification of CSTs may be more complex than initially conceived. Consequently, our DAR analysis based on Doyle et al. (2018) may be limited by the datasets of postpartum women, and the DAR parameters of the vaginal microbiomes of other CST women are likely different from the results reported here. Further studies should be performed to clarify this important issue.

The major findings in this study can be summarized using four profiles: i. e., DAR profile, characterizing the change (scaling) in diversity heterogeneity across individuals; PDO profile, characterizing the pair-wise similarity (overlap) between individuals; MAD profile, characterizing the maximal accrual diversity in a population; and RIP profile, characterizing the ratio of individual-level diversity to population-level diversity. Theoretically, the four profiles can together summarize the essential characteristics of the spatial distribution of vaginal microbial diversity and offer effective tools to sketch out the biogeographic maps of the human vaginal microbiome. Practically, they are essentially quantitative metrics of diversity heterogeneity across individuals from different dimensions (diversity scaling, pair-wise similarity in diversity, maximal accrual diversity, ratio of individual to population diversity). These multidimensional metrics could provide more comprehensive tools for understanding the implications of vaginal microbial diversity to women's health, including the DSR hypothesis for BV etiology (Ma et al., 2012; Ma & Ellison 2018, 2019; Sobel, 1999). In addition, the quantitative models of the four profiles obtained here could be harnessed to assess and predict microbiome diversity changes at the population scale and are of potential significance for evaluating women's health associated with

vaginal microbiomes.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Z.S.M. designed the study and wrote the paper. W.L. performed the data analysis and interpretation. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We are deeply indebted to Prof. Yong-Gang Yao for his advice and review of our manuscript. We are also deeply grateful to Dr. Ronan Doyle, Great Ormond Street Hospital, NHS Foundation Trust, United Kingdom, for his assistance in re-analyzing the raw sequencing reads for this study.

Wendy Li^{1,3}, Zhan-Shan (Sam) Ma^{1,2,3,*}

¹Computational Biology and Medical Ecology Lab, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming Yunnan 650223, China

²Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming Yunnan 650223, China

³Kunming College of Life Sciences, University of Chinese Academy of Sciences, Kunming Yunnan 650223, China

*Corresponding author, E-mail: ma@mail.kiz.ac.cn

REFERENCES

Bell T, Ager D, Song JI, Newman JA, Thompson IP, Lilley AK, van der Gast CJ. 2005. Larger islands house more bacterial taxa. *Science*, **308**(5730): 1884.

Brotman RM, Shardell MD, Gajer P, Fadrosh D, Chang K, Silver M, Viscidi RP, Burke AE, Ravel J, Gravitt PE. 2014. Association between the vaginal microbiota, menopause status and signs of vulvovaginal atrophy. *Menopause*, **21**(5): 450–458.

BuvéA, Jespers V, Crucitti T, Fichorova RN. 2014. The vaginal microbiota and susceptibility to HIV. *AIDS*, **28**(16): 2333–2344.

Chao A, Chiu C H, Hsieh TC. 2012. Proposing a resolution to debates on diversity partitioning. *Ecology*, **93**(9): 2037–2051.

Chao A, Chiu CH, Jost L. 2014. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Reviews of Ecology, Evolution, and Systematics*, **45**: 297–324.

Chu DM, Ma J, Prince AL, Antony KM, Seferovic MD, Aagaard KM. 2017. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nature Medicine*, **23**(3): 314–326.

Cox LM, Yamanishi S, Sohn J, Alekseyenko AV, Leung JM, Cho I, Kim SG, Li H, Gao Z, Mahana D, Rodriguez JGZ, Rogers AB, Robine N, Loke P,

Blaser MJ. et al. 2014. Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell*, **158**(4): 705–721.

DiGuilio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, Sun CL, Gotsman DSA, Wong RJ, Shaw G, Stevenson DK, Holmes SP, Relman DA. 2015. Temporal and spatial variation of the human microbiota during pregnancy. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(35): 11060–11065.

Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R. 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(26): 11971–11975.

Doyle R, Gondwe A, Fan Y M, Maleta K, Ashorn P, Klein N, Harris K. 2018. A *Lactobacillus*-deficient vaginal microbiota dominates postpartum women in rural Malawi. *Applied and Environmental Microbiology*, **84**(6): e02150–17.

Ellison AM. 2010. Partitioning diversity. *Ecology*, **91**(7): 1962–1963.

Freitas AC, Bocking A, Hill JE, Money DM, VOGUE Research Group. 2018. Increased richness and diversity of the vaginal microbiota and spontaneous preterm birth. *Microbiome*, **6**: 117.

Freitas AC, Chaban B, Bocking A, Rocco M, Yang S, Hill JE, Money DM, VOGUE Research Group. 2017. The vaginal microbiome of pregnant women is less rich and diverse, with lower prevalence of Mollicutes, compared to non-pregnant women. *Scientific Reports*, **7**: 1–16.

Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UM, Zhong X, Koenig SS, Fu L, Ma ZS, Zhou X, Abdo Z, Forney LJ, Ravel J. 2012. Temporal dynamics of the human vaginal microbiota. *Science Translational Medicine*, **4**(132): 132ra52.

Gotelli NJ, Chao A. 2013. Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. In: Levin SA. Encyclopedia of Biodiversity (Vol 5). Waltham, MA; OA academic Press, 195–211.

Green J, Bohannan BJM. 2006. Spatial scaling of microbial biodiversity. *TRENDS in Ecology and Evolution*, **21**(9): 501–507.

Hickey RJ, Zhou X, Settles ML, Erb J, Malone K, Hansmann MA, Shew ML, Van Der Pol B, Fortenberry JD, Forney LJ. 2015. Vaginal microbiota of adolescent girls prior to the onset of menarche resemble those of reproductive-age women. *mBio*, **6**(2): e00097-15. doi: 10.1128/mBio.00097-15.

Hill MO. 1973. Diversity and evenness: A unifying notation and its consequences. *Ecology*, **54**(2): 427–432.

Horner-Devin M C, Lage M, Hughes JB, Bohannan BJM. 2004. A taxa-area relationship for bacteria. *Nature*, **432**: 750–753.

Jost L. 2007. Partitioning diversity into independent alpha and beta components. *Ecology*, **88**(10): 2427–2439.

Ma B, Forney LJ, Ravel J. 2012. Vaginal microbiome: rethinking health and disease. *Annual Review of Microbiology*, **66**: 371–389.

Ma ZS. 2018a. DAR (diversity–area relationship): Extending classic SAR (species – area relationship) for biodiversity and biogeography analyses. *Ecology and Evolution*, **8**(20): 10023–10038.

Ma ZS. 2018b. Diversity time-period and diversity-time-area relationships exemplified by the human microbiome. *Scientific Reports*, **8**: 7214.

Ma ZS. 2018c. Sketching the human microbiome biogeography with DAR (Diversity Area Relationship) profiles. *Microbial Ecology*, doi: 10.1007/s00248-018-1245-6.

Ma ZS. 2019. A new DTAR (diversity – time – area relationship) model demonstrated with the indoor microbiome. *Journal of Biogeography*, doi: 10.1111/jbi.13636.

Ma ZS, Ellison AM. 2018. A unified concept of dominance applicable at both community and species scales. *Ecosphere*, **9**(11): e02477.

Ma ZS, Ellison AM. 2019. Dominance network analysis provides a new framework for studying the diversity-stability relationship. *Ecological Monographs*, **89**(2): e01358.

Ma ZS, Li LW. 2018. Semen microbiome biogeography: an analysis based on a Chinese population study. *Frontiers in Microbiology*, **9**: 3333.

Ma ZS, Li LW. 2017. Quantifying the human vaginal community state types (CSTs) with the species specificity index. *PeerJ*, **5**: e3366.

MacArthur RH, Wilson EO. 1967. The Theory of Islands Biogeography. Princeton: Princeton University Press.

Macklaim J M, Fernandes A D, Di Bella J M, Hammond JA, Reid G, Gloor GB. 2013. Comparative meta-RNA-seq of the vaginal microbiota and differential expression by *Lactobacillus* iners in health and dysbiosis. *Microbiome*, **1**: 12.

Muhleisen AL, Herbst-Kralovetz MM. 2016. Menopause and the vaginal microbiome. *Maturitas*, **91**: 42–50.

Noguera AM, Arita HT, Escalante AE, Forney LJ, García-Olivá F, Souza V. 2005. Microbial macroecology: highly structured prokaryotic soil assemblages in a tropical deciduous forest. *Global Ecology and Biogeography*, **14**(3): 241–248.

Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, Glickman JN, Siebert R, Baron RM, Kasper DL, Blumberg RS. 2012. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science*, **336**(6080): 489–493.

Peay KG, Bruns TD, Kennedy PG, Bergemann SE, Garbelotto M. 2007. A strong species–area relationship for eukaryotic soil microbes: island size matters for ectomycorrhizal fungi. *Ecology Letters*, **10**(6): 470–480.

Petrova MI, van de BM, Balzarini J, Vanderleyden J, Lebeur S. 2013. Vaginal microbiota and its role in HIV transmission and infection. *FEMS Microbiology Reviews*, **37**(5): 762–792.

Ralph SG, Rutherford AJ, Wilson JD. 1999. Influence of bacterial vaginosis on conception and miscarriage in the first trimester: cohort study. *British Medical Journal*, **319**(7204): 220–223.

Ravel J, Brotman RM, Gajer P, Ma B, Nandy M, Fadrosh DW, Sakamoto J, Koenig SK, Fu L, Zhou X, Hickey RJ, Schwebke JR, Forney LJ. 2013. Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. *Microbiome*, **1**: 29.

Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. 2011. Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America*, **108**: 4680–4687.

Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Bieda J, Chaemsathong P, Miranda J, Chaiworapongsa T, Ravel J. 2014a. The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. *Microbiome*, **2**: 1–15.

Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Nikita L, Galuppi

M, Lamont RF, Chaemsathong P, Miranda J, Chaiworapongsa T, Ravel J. 2014b. The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome*, **2**: 4.

Rutayisire E, Huang K, Liu Y, Tao F. 2016. The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review. *BMC Gastroenterol*, **16**(1): 86.

Sevelsted A, Stokholm J, Bønnelykke K, Bisgaard H. 2015. Cesarean section and chronic immune disorders. *Pediatrics*, **135**(1): e92–8.

Sobel JD. 1999. Is there a protective role for vaginal flora?. *Current Infectious Disease Reports*, **1**: 379–383.

Stout MJ, Zhou YJ, Wylie KM, Tarr PI, Macones GA, Tuuli MG. 2017. Early pregnancy vaginal microbiome trends and preterm birth. *American Journal of Obstetrics and Gynecology*, **217**(3): 356.e1–356.e18.

Triantis KA, Guilhaumon F, Whittaker RJ. 2012. The island species–area relationship: biology and statistics. *Journal of Biogeography*, **39**(2): 215–231.

Várbíró G, Görgényi J, Tóthmérész B, Padisák J, Hajnal É, Borics G. 2017. Functional redundancy modifies species–area relationship for freshwater phytoplankton. *Ecology and Evolution*, **7**(23): 9905–9913.

Watson HC. 1835. Remarks on the Geographical Distribution of British Plants. London: Longman Press.

Whittaker RJ, Triantis KA. 2012. The species – area relationship: an exploration of that ‘most general, yet protean pattern’. *Journal of Biogeography*, **39**(4): 623–626.

Younes JA, Lievens E, Hummelen R, van der Westen R, Reid G, Petrova MI. 2018. Women and their microbes: the unexpected friendship. *Trends in Microbiology*, **26**(1): 16–32.